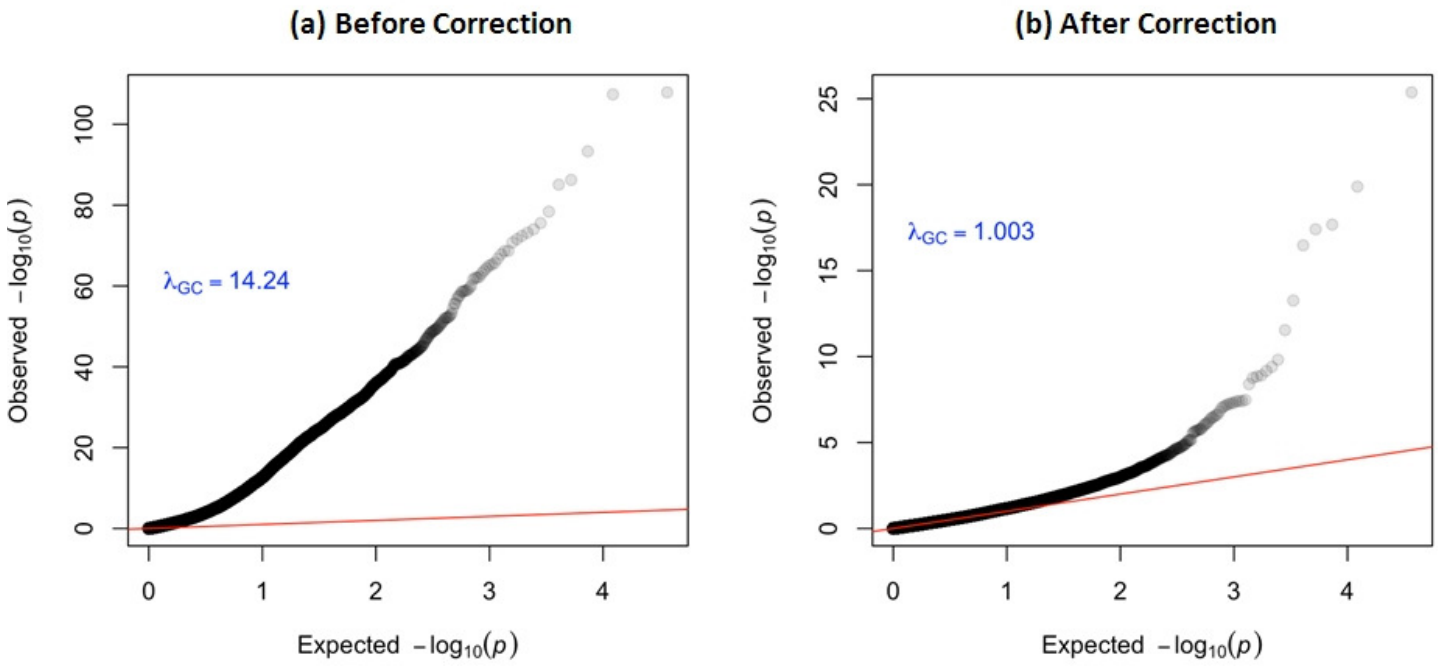
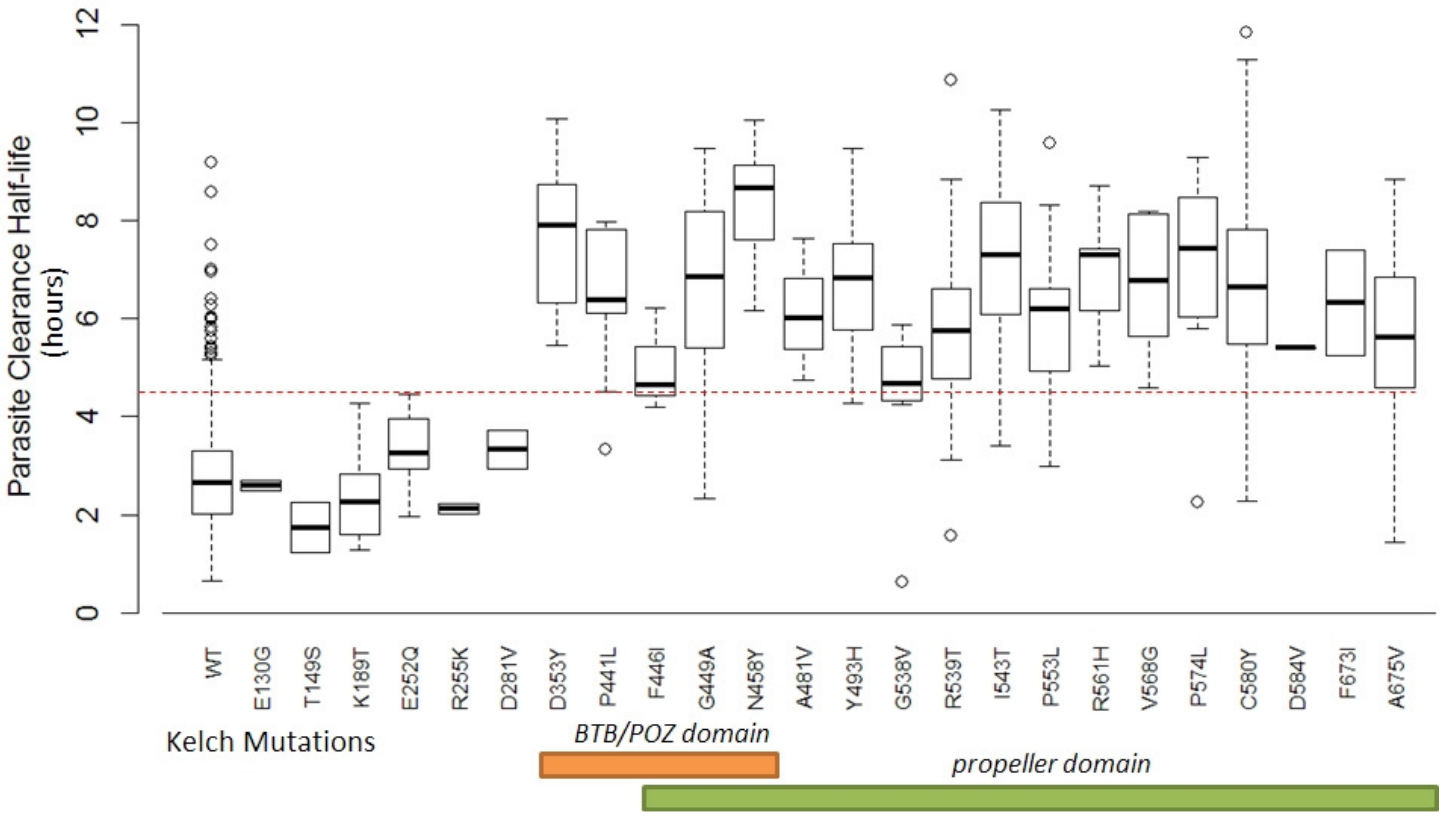


Supplementary Figure 1

Effect of correction for population structure in the genome-wide association study.

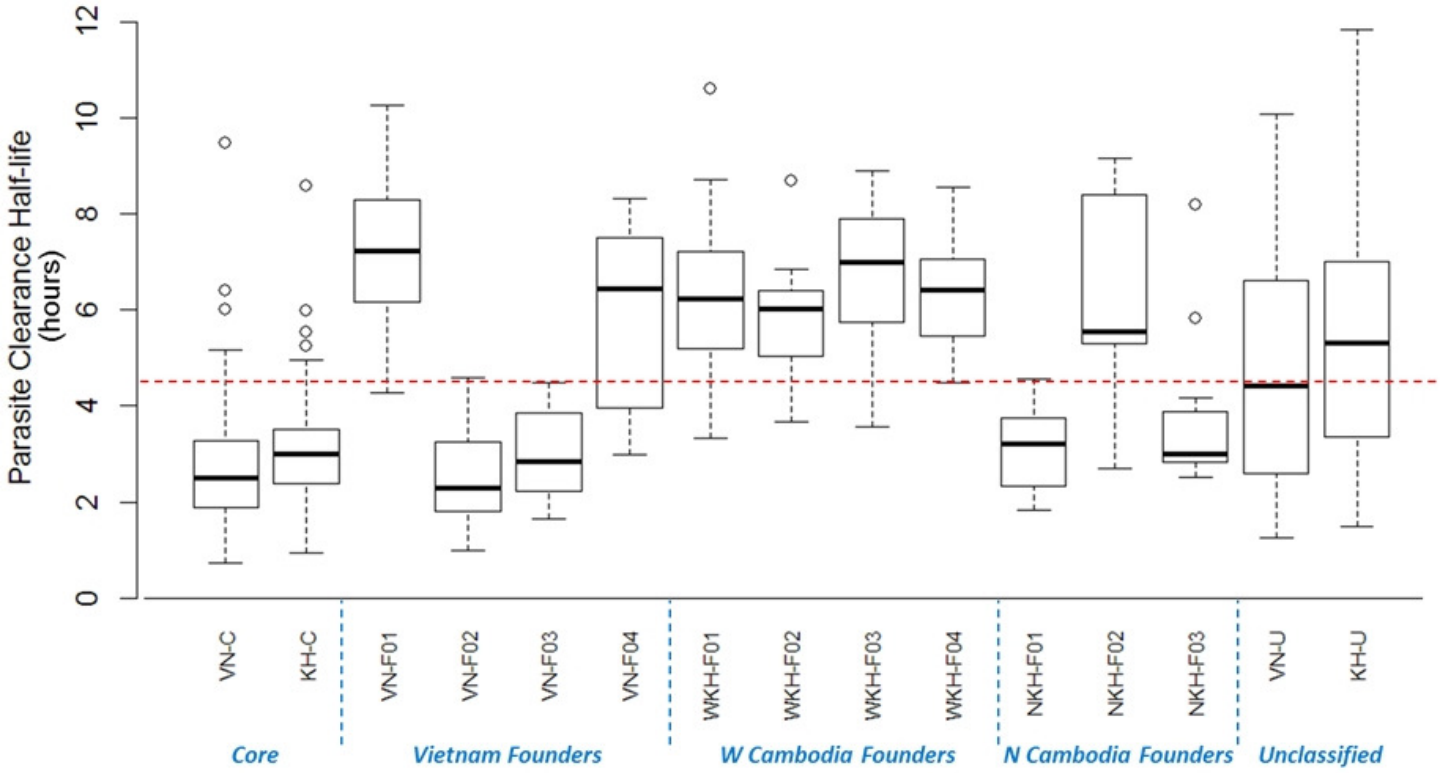
Quantile-quantile (QQ) plots showing the effect of the genetic relatedness between samples on the GWAS, before and after correction. The expected distribution of *P* values (*x* axis) is compared to that of observed values (*y* axis) across all tested SNPs, and the genomic inflation factor λ_{GC} is indicated. (**a**) *P*-value distributions before correction, showing very high inflation due to population structure. (**b**) *P*-value distributions after using a genetic relatedness matrix as a covariate in the analysis, a correction resulting in a significant decrease in the inflation factor. Note that the *y*-axis scales are different in the two plots.



Supplementary Figure 2

Distributions of parasite clearance half-life ($PCt_{1/2}$) for different *kelch13* mutations.

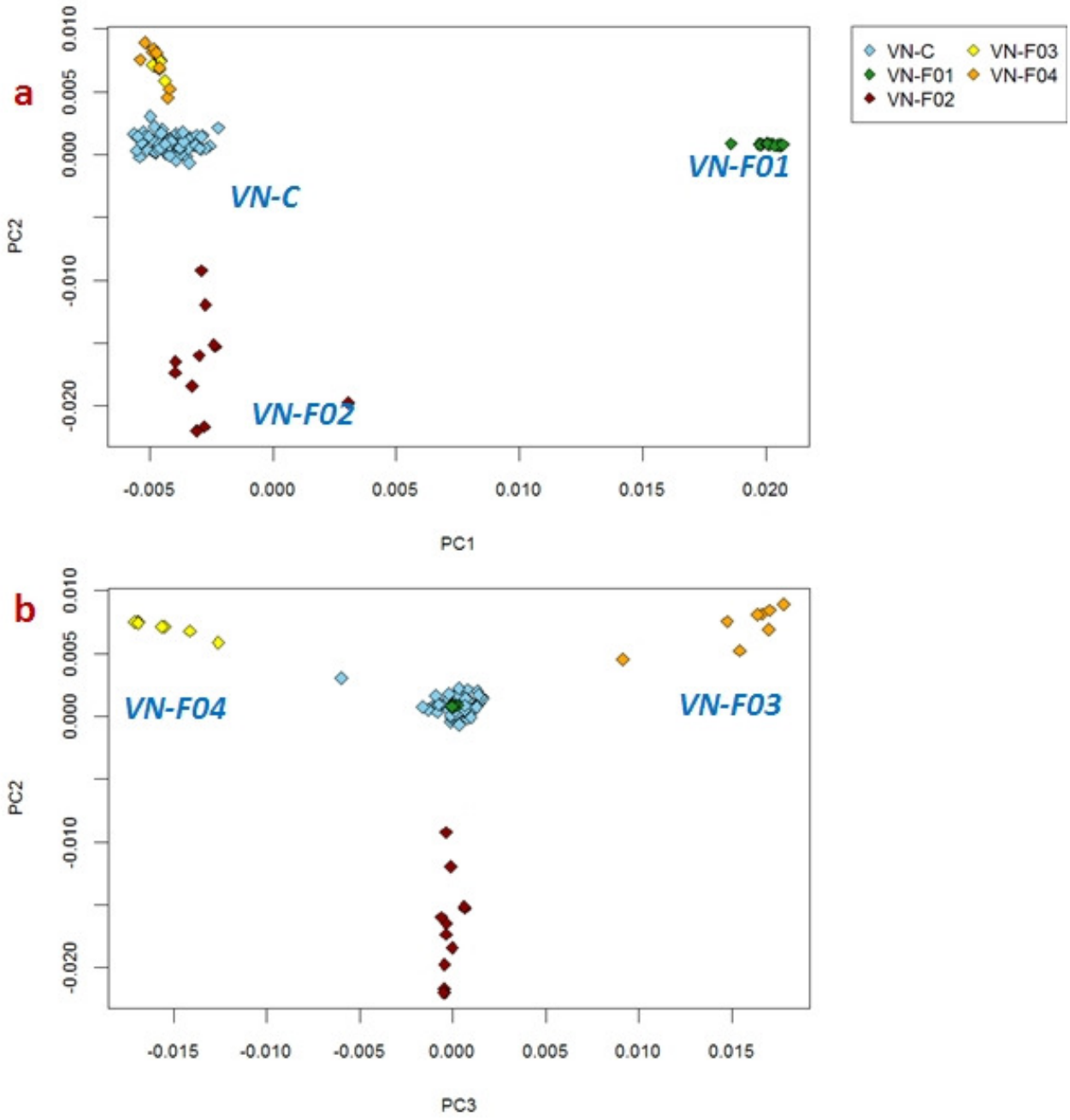
$PCt_{1/2}$ values for samples carrying 1 of 24 mutant *kelch13* alleles (n = 469) are compared to those carrying wild-type (WT) alleles (n = 630); samples with heterozygous or missing calls were excluded (n = 190). Alleles carried by a single sample are excluded from this diagram (n = 8). Plots show median, interquartile range, estimated 95% confidence interval and outlying values. A red dotted line shows $PCt_{1/2}$ = 4.5 h, a notional boundary of clinical artemisinin resistance. Below the x axis, colored bands identify those mutations occurring in either the BTB/BOZ domain or the propeller domain.



Supplementary Figure 3

Distributions of parasite clearance half-life in core and founder populations in Vietnam and Cambodia.

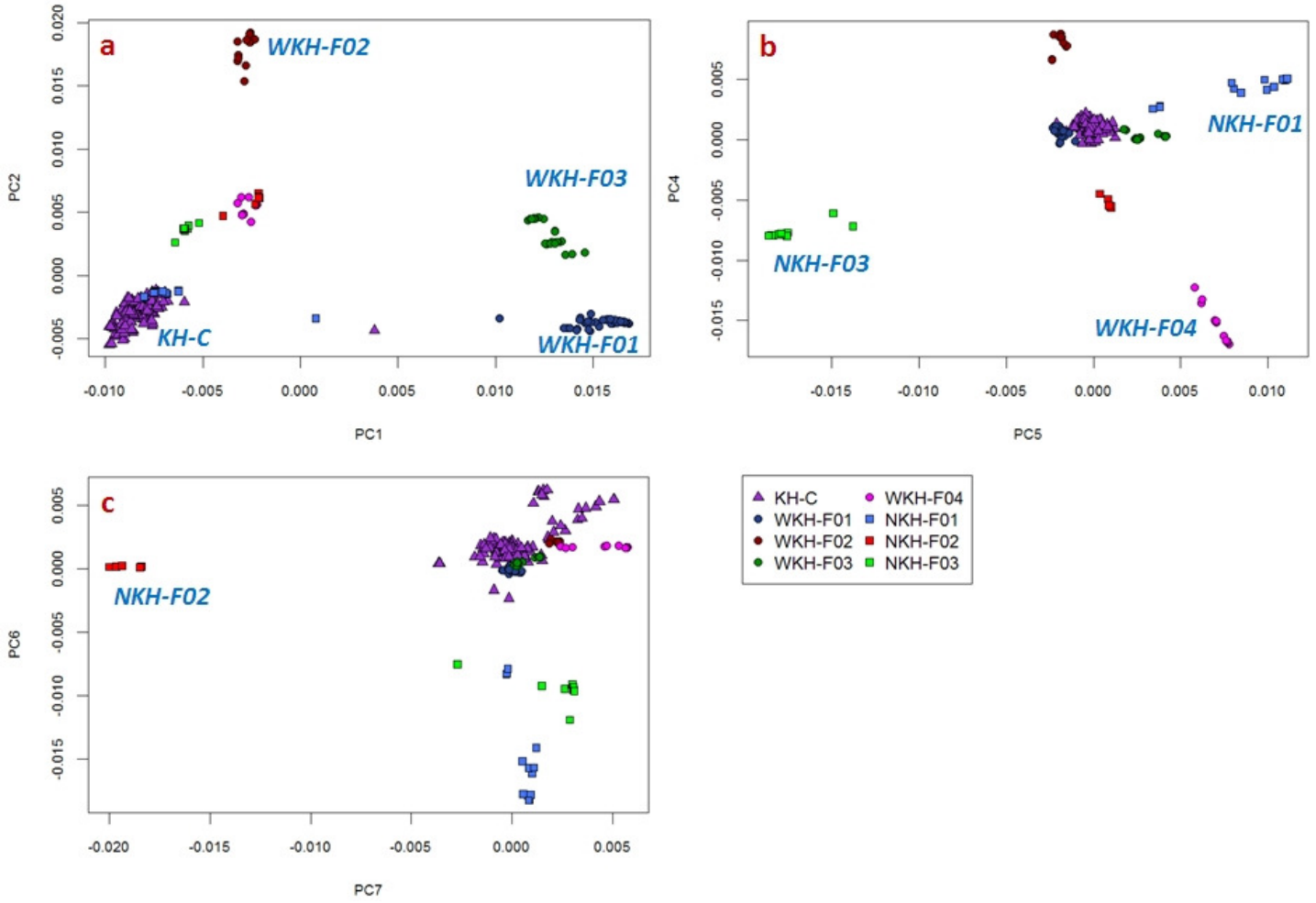
Distributions of parasite clearance half-life ($PCt_{1/2}$) in all parasite subpopulations identified in Vietnam and Cambodia (**Supplementary Tables 5–8**) are shown by box plots indicating median, interquartile range, estimated 95% confidence interval and outlying values. A red horizontal dotted line indicates $PCt_{1/2}$ = 4.5 h, a notional boundary between sensitive and resistant phenotypes. Two founder populations in Vietnam (VN-F02 and VN-F03) and two in northern Cambodia (NKH-F01 and NKH-F03) comprise artemisinin-sensitive parasites, whereas the remaining founder populations are artemisinin resistant.



Supplementary Figure 4

Correspondence of Vietnamese founder populations with PCoA outlier groups.

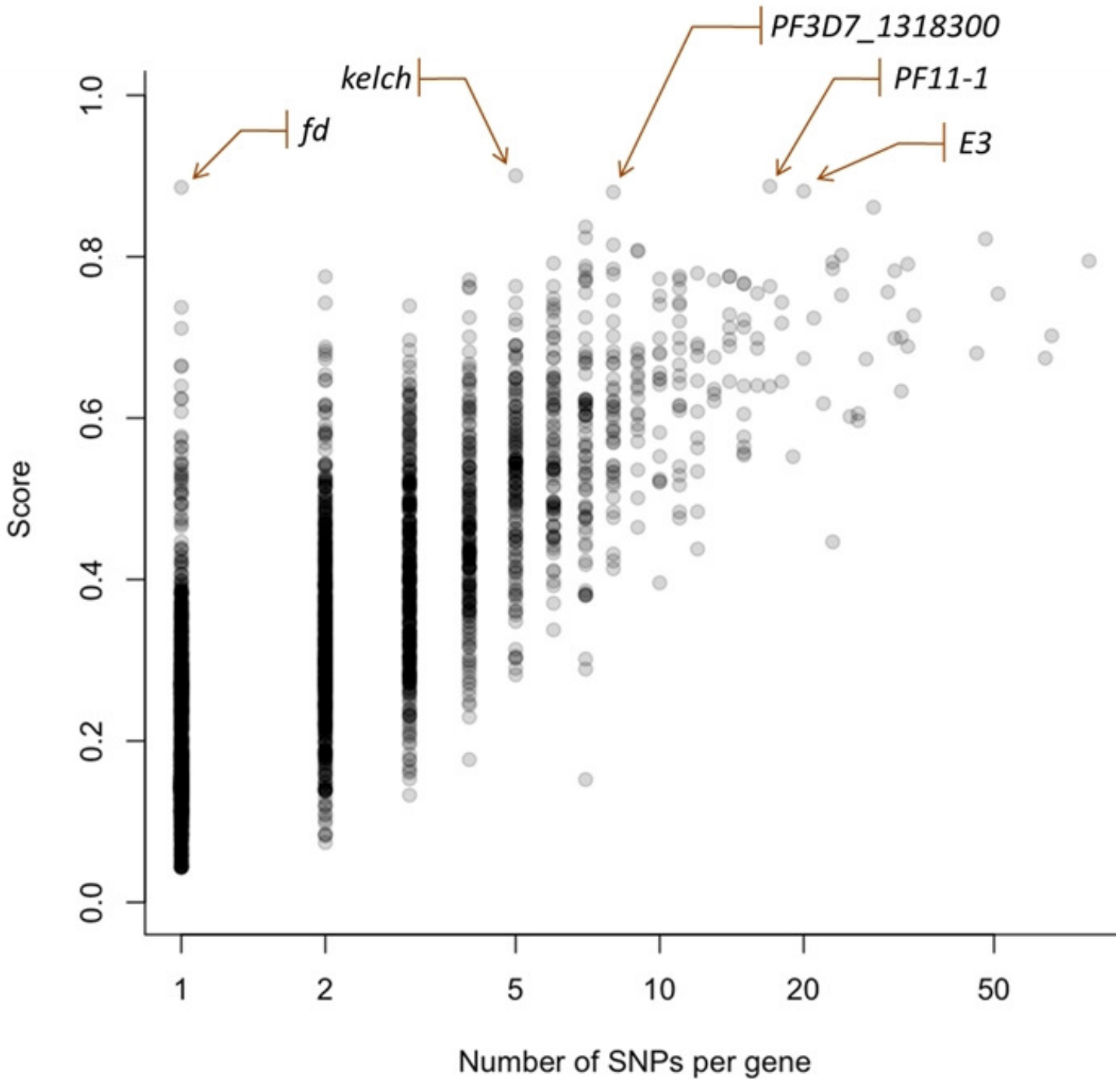
In this PCoA analysis, Vietnamese samples are colored according to their assigned population. The plots show different components for the same sample set: (**a**) PC1 versus PC2 and (**b**) PC2 versus PC3. Unclassified samples (VN-U) are omitted for clarity. There is a strong correspondence between the founder populations identified by ancestry analysis and the outlier groups emerging from these plots (indicated by blue labels).



Supplementary Figure 5

Correspondence of Cambodian founder populations with PCoA outlier groups.

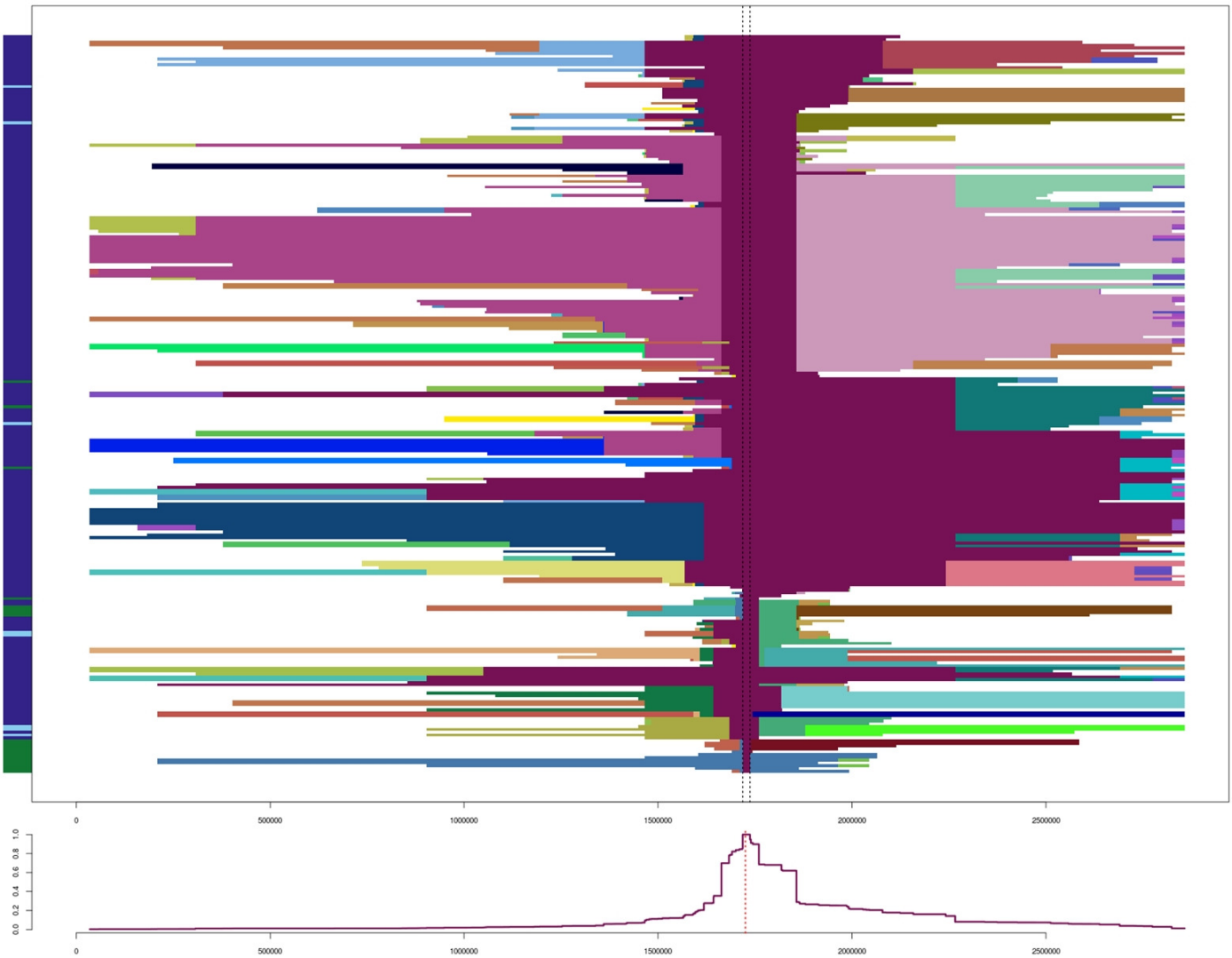
In this PCoA analysis, Cambodian samples are colored according to their assigned population. The plots show different components for the same sample set: (**a**) PC1 versus PC2, (**b**) PC4 versus PC5 and (**c**) PC6 versus PC7. Unclassified samples (KH-U) are omitted for clarity. There is a strong correspondence between the founder populations identified by ancestry analysis and the outlier groups emerging from these plots (indicated by blue labels).



Supplementary Figure 6

Distribution of gene heteroallelic scores of founder population differentiation.

For each gene, the F_{ST} value between each founder population and its corresponding core population was estimated, and the gene was assigned a score equal to the mean of all F_{ST} estimates. Scores in 2,518 genes analyzed are plotted against the number of SNPs with $F_{ST} \geq 0.3$ in the gene, showing that scores tend to rise with SNP counts. Points corresponding to the top five genes in this analysis are labeled.



Supplementary Figure 7

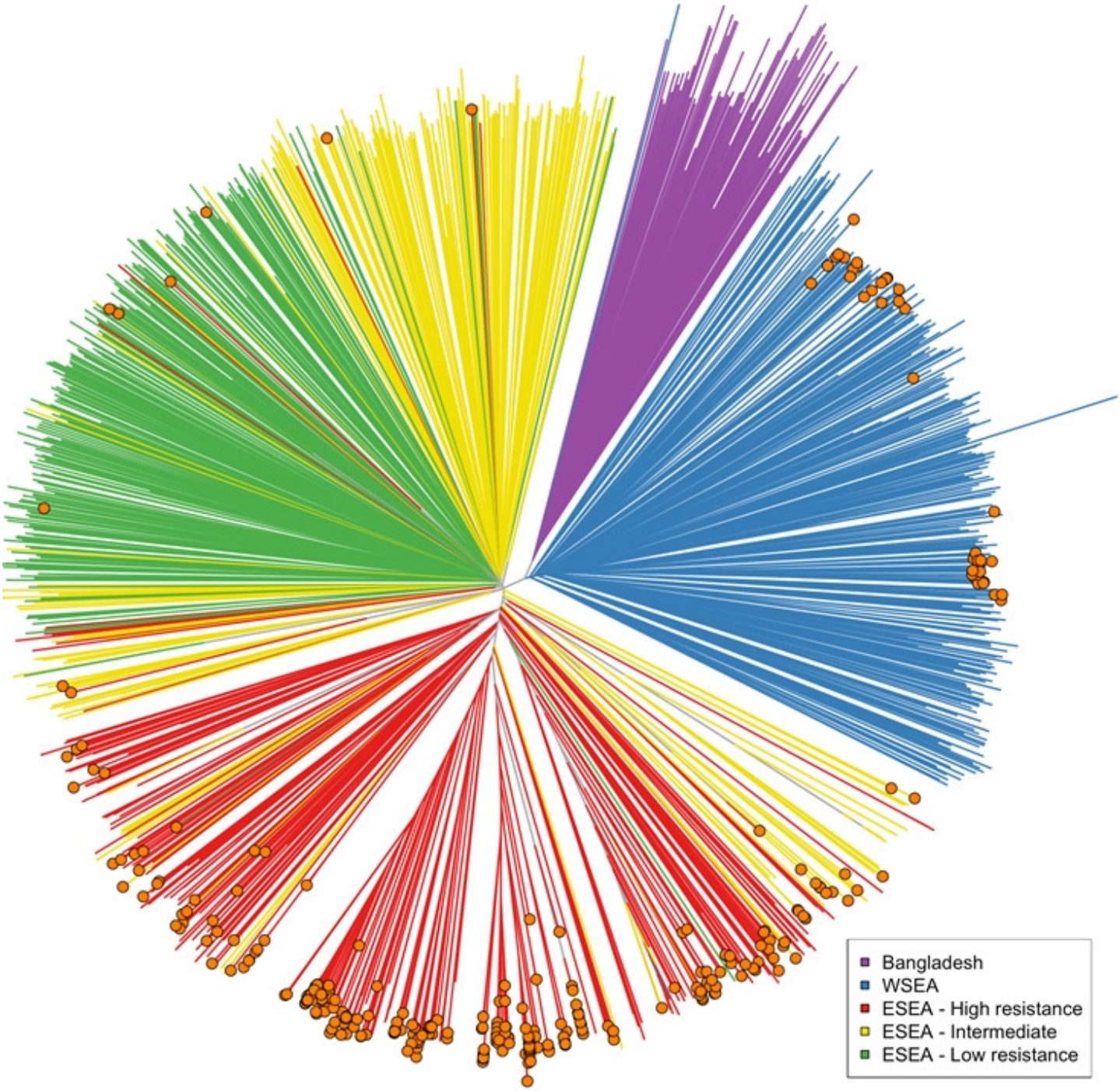
Decay of haplotypes on chromosome 13 in samples carrying the C580Y *kelch13* mutation.

In the top panel, each horizontal line represents a haplotype, and each color is used to signify a shared haplotype (i.e., carrying identical genotypes). On the left-hand side, colored bars indicate the country of origin of the sample (blue, Cambodia; cyan, Vietnam; green, Thailand). Two vertical dashed lines mark a 17-kb core region that is identical in all samples. Haplotype homozygosity decay is reported in the bottom panel; a red dashed line marks the position of the C580Y allele. In the upper part of the plot, clusters in ESEA present extremely long haplotypes (sometimes chromosome-wide), consistent with rapid population expansions. These clusters share a sizeable core haplotype, suggesting a common origin of their C580Y mutations. By comparison, the haplotype shared with C580Y mutants from western Thailand (bottom of the plot) is far shorter, suggesting an independent origin of the mutation in the two regions.

Supplementary Figure 8

Pairwise longest common haplotype length tree for samples containing *kelch13* mutations, tips colored by mutation and country of origin.

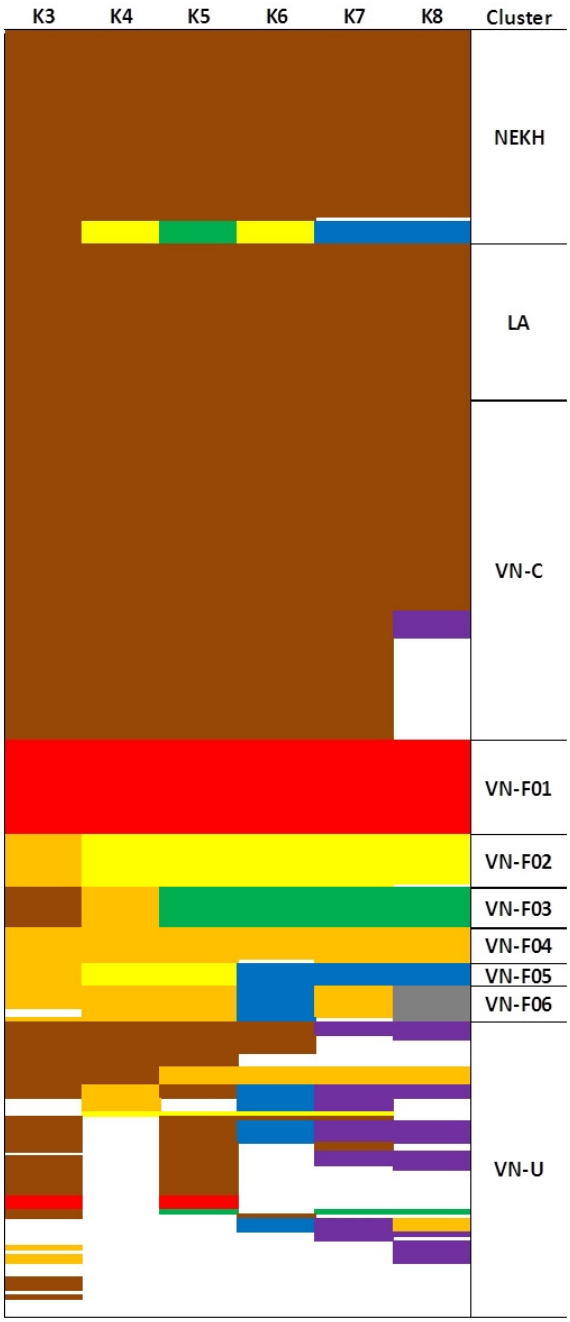
Tips represent samples and have been colored by mutation and country of origin (KH, Cambodia; VN, Vietnam; LA, Laos; TH, western Thailand; ET, eastern Thailand; MM, Myanmar). Internal branches have been colored by the most frequent mutation present in the subtended subtrees. The bars at the bottom offer visual aid for tracking how different mutations cluster and segregate across the tree.



Supplementary Figure 9

Neighbour-joining tree showing samples carrying the *kelch13* C580Y mutation.

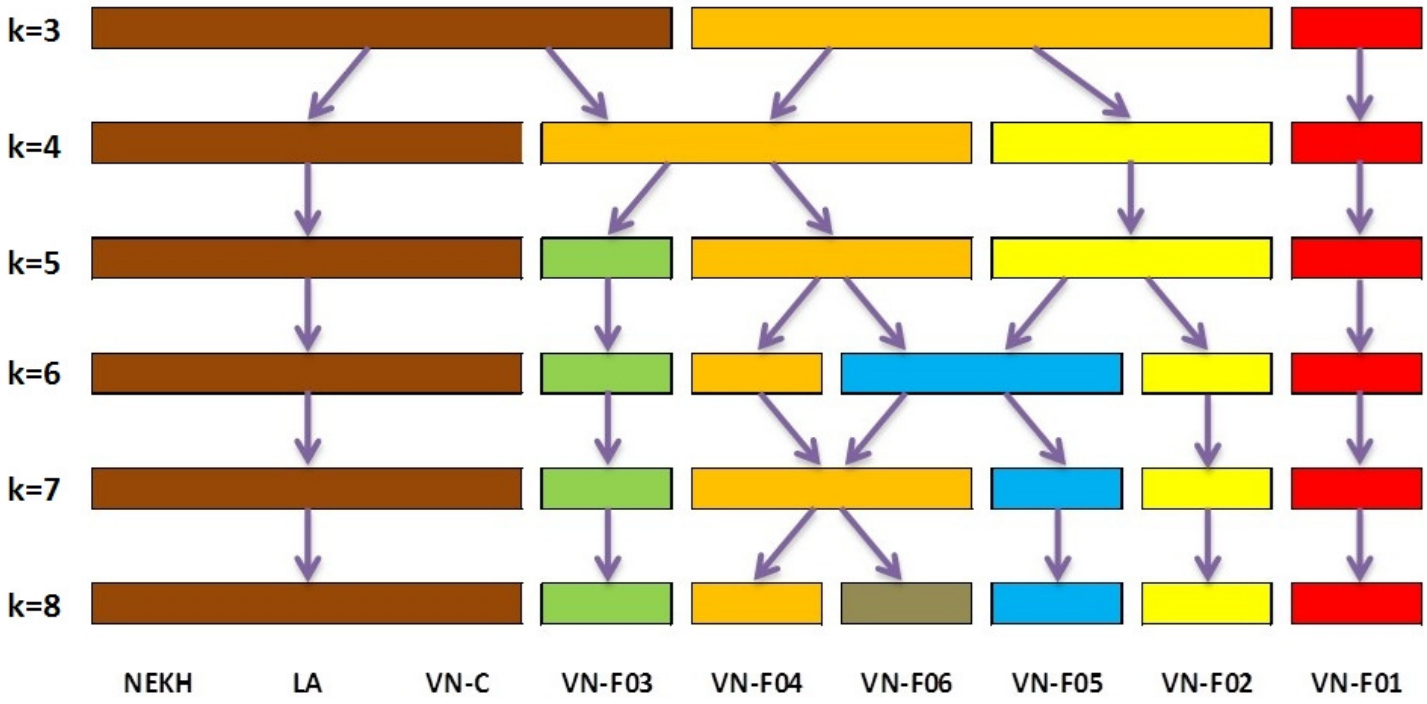
Analogous to **Figure 3a**, this neighbour-joining tree shows samples colored by their population compartments. Orange tip symbols denote samples carrying the *kelch13* C580Y mutation.



Supplementary Figure 10

Ancestry analysis of Vietnamese populations.

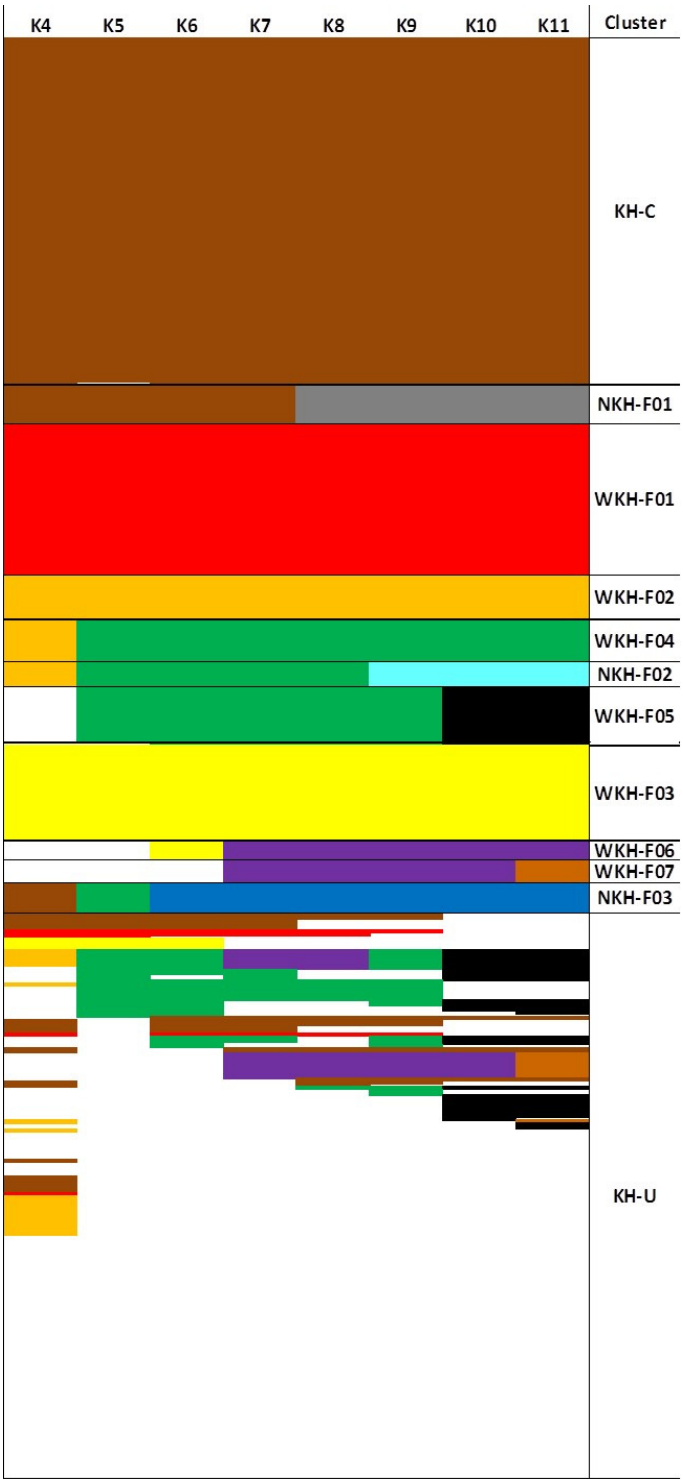
The figure shows the ancestry assignments for each sample, under six different hypotheses of ancestor numbers (k = 3 to 8). Each row corresponds to a sample, and the colors in the first six columns indicate group assignments under each hypothesis, as detailed in the **Supplementary Note**; each color represents one group, corresponding to one of the k ancestries, while unclassified samples are shown in white. Vietnamese samples that grouped together across all hypotheses (one mismatch allowed) were grouped into clusters if these clusters comprised =5 samples (labeled VN-C and VN-F01 to VN-F06); other Vietnamese samples were assigned to an 'unclassified' group (VN-U). The VN-C cluster was identified by similarity with the Cambodian and Laotian populations (NEKH and LA, respectively).



Supplementary Figure 11

Cluster stability analysis of Vietnamese populations.

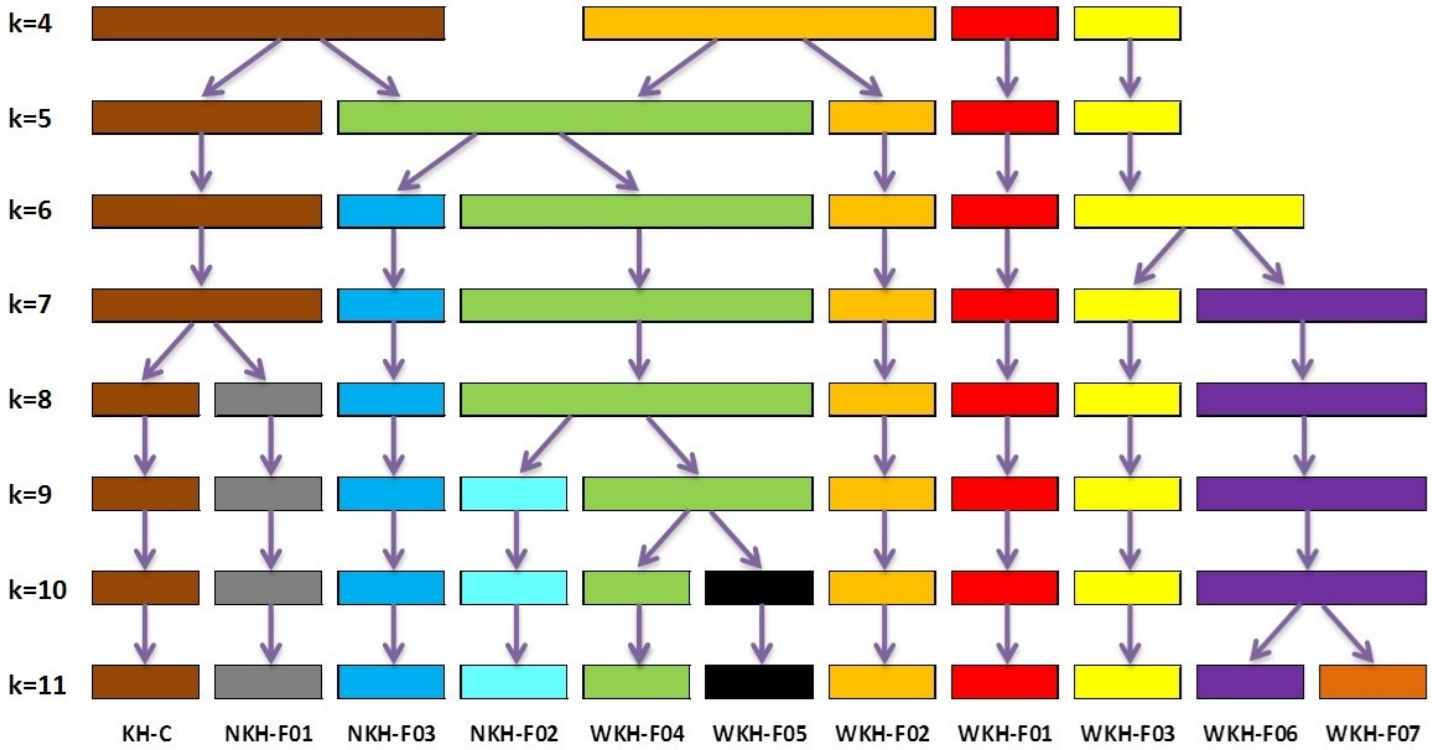
The diagram shows the relationships between the clusters identified for different hypotheses, using the sample assignments shown in **Supplementary Figure 10**. Non-core populations separate at different values of k, e.g. at k = 3 (VN-F01) and k = 4 (VN-F02, VN-F03 and VN-F04). Cluster VN-F01 is remarkably stable, reflecting a high degree of differentiation, and the classifications of clusters VN-F02, VN-F03 and VN-F04 also appear consistent across different values of k. Conversely, clusters VN-F05 and VN-F06 appear to aggregate with different groups as k increases, suggesting that they may not represent a stable subpopulation. This effect is possibly aggravated by the very small size of these clusters.



Supplementary Figure 12

Ancestry analysis of Cambodian populations.

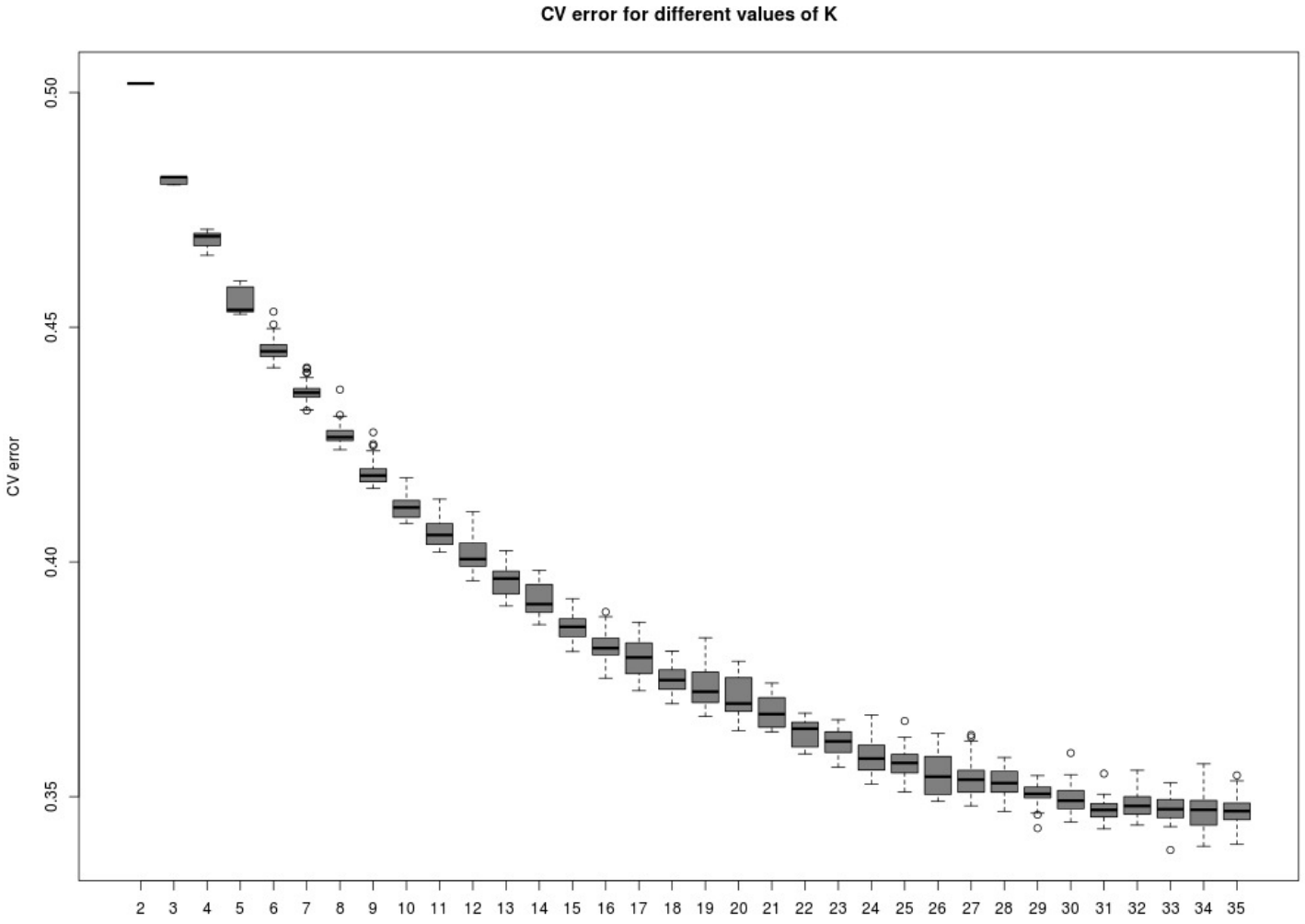
The figure shows the ancestry assignments for each sample, under eight different hypotheses of ancestor numbers (k = 4 to 8). The method and coloring scheme are the same as for **Supplementary Figure 10**.



Supplementary Figure 13

Cluster stability analysis of Cambodian populations.

The diagram shows the relationships between the clusters identified for different hypotheses, using the sample assignments shown in **Supplementary Figure 12**. The smaller clusters KH-F06 and KH-F07 were deemed unstable and discarded, as they comprise samples that were unclassified for k = 6, as well as samples previously grouped with KH-F03.



Supplementary Figure 14

Cross-validation (CV) error distribution for Cambodian populations.

This figure shows the CV error distribution of 50 ADMIXTURE runs with different random seeds for each value of K. As K increases, the CV error reaches a plateau with no clear minimum; solutions show marginal improvements, while error variance grows.

Supplementary Tables

Supplementary Table 1 – Additional samples whose genomes were included in population structure studies, but not in the genome-wide association study (GWAS). These samples were contributed by various studies participating in the MalariaGEN *Plasmodium falciparum* Community Project (for further details, see http://www.malariagen.net/projects/parasite/pf). Contributors are abbreviated as follows: ARC3 – Artemisinin Resistance Confirmation, Characterization and Containment;[1,2] ARCE – Containment of Artemisinin Tolerant Malaria Parasites in Southeast Asia Project;[3-5] MORU – Mahidol-Oxford Research Unit, Bangkok, Thailand;[6] SMRU – Shoklo Malaria Research Unit, Mae Sot, Thailand;[7] Texas Biomed – Texas Biomedical Research Institute, San Antonio, TX, USA;[8] UMD/HHMI – University of Maryland School of Medicine, Howard Hughes Medical Institute, Baltimore, MD, USA;[1] OUCRU – Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam;[3] NIAID/NIH – National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA;[9,10] TRAC – Tracking Resistance to Artemisinin Collaboration.[11] Abbreviations used in this paper to indicate the respective geographical regions are shown in the "Code" column.

Contributor	Years	Country (Region)	Code	Locations	Samples
MORU/SMRU, Texas Biomed	2001-08	Thailand (West)	WTH	Mae Sot	108
MORU	2007	Cambodia (West)	WKH	Pailin	3
ARC3	2008-09	Bangladesh	BD	Bandarban	29
		Cambodia (West)	WKH	Pailin	35
				Tasanh	49
		Thailand (West)	WTH	Mae Sot	3
UMD/HHMI	2009	Cambodia (West)	WKH	Pailin	15
OUCRU	2009-10	Vietnam	VN	Binh Phuoc	19
ARCE	2010-11	Laos	LA	Xepon	36
		Myanmar (South)	SMM	Kawthaung	50
		Vietnam	VN	Binh Phuoc	108
NIAID/NIH (*)	2008-10	Cambodia (West)	WKH	Pursat	67
		Cambodia (Northeast)	NEKH	Ratanakiri	4
TRAC (*)	2011-13	Myanmar (Central)	CMM	Bago Division	1
		Laos	LA	Attapeu	8
		Thailand (West)	WTH	Mae Sot	3
		Cambodia (West)	WKH	Pursat	3
				Total	**541**

(*) Samples not included in the GWAS because phenotypes were not available.

Supplementary Table 2 – SNPs showing the most significant associations in the genome-wide association study (GWAS). This table includes all SNPs with $P \leq 10^{-5}$ ($n = 48$), ordered by their genomic coordinates. SNPs significant with a Bonferroni-corrected threshold ($P \leq 10^{-7}$) are shown in bold type. For each SNP, we show chromosome number; nucleotide position; rank of the SNP in the GWAS; gene id and description; whether the SNP is non-synonymous or synonymous; the mutation name; and the association *P*-value.

Chr	Pos	Rank	Gene Id	Gene Description	N/S	Mutation	*P*
5	852479	47	PF3D7_0520800	conserved Plasmodium protein, unknown function	N	S1183N	7.56E-06
	855646	37	PF3D7_0520800	conserved Plasmodium protein, unknown function	S	127N	1.70E-06
7	**405362**	**13**	**PF3D7_0709000**	**chloroquine resistance transporter (CRT)**	**N**	**N326S**	**1.73E-09**
	405600	**10**	**PF3D7_0709000**	**chloroquine resistance transporter (CRT)**	**N**	**I356T**	**6.85E-10**
	896660	**9**	**PF3D7_0720700**	**phosphoinositide-binding protein, putative**	**N**	**C1484F**	**3.95E-10**
8	392003	25	PF3D7_0807600	conserved Plasmodium protein, unknown function	N	N578D	1.56E-07
	481897	48	PF3D7_0809600	peptidase family C50, putative	N	K5162N	9.01E-06
	497184	43	PF3D7_0809600	peptidase family C50, putative	N	I67L	2.96E-06
	789468	41	PF3D7_0817300	asparagine-rich antigen	N	K50R	2.38E-06
	1392729	36	PF3D7_0832200.1	Plasmodium exported protein (PHISTa-like), unknown function	N	T152I	1.66E-06
9	973634	39	PF3D7_0924000	patatin-like phospholipase, putative	N	K363E	2.04E-06
	973647	33	PF3D7_0924000	patatin-like phospholipase, putative	N	G367E	8.32E-07
	973686	35	PF3D7_0924000	patatin-like phospholipase, putative	N	I380N	1.22E-06
	1055456	34	PF3D7_0926100	protein kinase, putative	N	D787E	9.32E-07
10	490648	28	PF3D7_1012700	protein phosphatase, putative	N	Y1133N	3.24E-07
	490720	**23**	**PF3D7_1012700**	**protein phosphatase, putative**	**N**	**V1157L**	**7.93E-08**
	497461	31	PF3D7_1012900	conserved Plasmodium protein, unknown function	N	T38I	5.89E-07
12	739074	46	PF3D7_1218800	conserved Plasmodium protein, unknown function (PSOP17)	N	R194K	7.52E-06
13	**748395**	**5**	**PF3D7_1318100**	**ferredoxin, putative**	**N**	**D193Y**	**3.37E-17**
	754133	**11**	**PF3D7_1318300**	**conserved Plasmodium protein, unknown function**	**N**	**T75I**	**1.24E-09**
	958469	**22**	**PF3D7_1322700**	**conserved Plasmodium protein, unknown function**	**N**	**T236I**	**6.88E-08**
	1700345	**20**	**PF3D7_1343100**	**conserved Plasmodium protein, unknown function**	**N**	**N96D**	**5.26E-08**
	1717359	**3**	**PF3D7_1343400**	**DNA helicase, putative**	**N**	**N821K**	**2.10E-18**
	1718288	**4**	**PF3D7_1343400**	**DNA helicase, putative**	**N**	**N1131I**	**3.94E-18**
	1725259	**1**	**PF3D7_1343700**	**kelch protein, putative**	**N**	**C580Y**	**4.24E-26**

Chr	Pos	Rank	Gene Id	Gene Description	N/S	Mutation	*P*
	1739315	**15**	**PF3D7_1343800**	**conserved Plasmodium protein, unknown function**	**S**	**4724N**	**3.34E-08**
	1820881	**24**	**PF3D7_1345400**	**conserved Plasmodium protein, unknown function**	**S**	**140T**	**9.65E-08**
	1821294	44	PF3D7_1345400	conserved Plasmodium protein, unknown function	N	D3H	7.09E-06
	1862741	**17**	**PF3D7_1346400**	**conserved Plasmodium protein, unknown function**	**N**	**D3282H**	**3.88E-08**
	1913206	30	PF3D7_1347900	conserved Plasmodium protein, unknown function	N	T1814A	4.17E-07
	1988624	29	PF3D7_1349500	conserved Plasmodium protein, unknown function	N	D2331N	3.51E-07
	1991685	**14**	**PF3D7_1349500**	**conserved Plasmodium protein, unknown function**	**S**	**1310K**	**4.03E-09**
	2015894	42	PF3D7_1350500	conserved Plasmodium protein, unknown function	N	D970N	2.43E-06
	2028330	**21**	**PF3D7_1350900**	**transcription factor with AP2 domain(s) (ApiAP2)**	**S**	**103L**	**5.69E-08**
	2600019	26	PF3D7_1364600	aldo-keto reductase, putative	S	421P	2.29E-07
	1481740	27	PF3D7_1436300	translocon component PTEX150 (PTEX150)	N	D655A	2.61E-07
	1956225	**8**	**PF3D7_1447900**	**multidrug resistance protein 2+(heavy metal transport family) (MDR2)**	**N**	**T484I**	**1.53E-10**
	1960049	40	PF3D7_1448000	U3 snoRNA-associated small subunit rRNA processing protein, putative	N	H1271R	2.06E-06
	2079623	**16**	**PF3D7_1450700**	**conserved Plasmodium protein, unknown function**	**N**	**T379I**	**3.86E-08**
	2081506	32	PF3D7_1450800	conserved Plasmodium protein, unknown function	N	G248R	7.26E-07
	2082099	**19**	**PF3D7_1450800**	**conserved Plasmodium protein, unknown function**	**N**	**S50N**	**5.00E-08**
14	**2086413**	**18**	**PF3D7_1451000**	**conserved Plasmodium protein, unknown function**	**N**	**N369K**	**4.02E-08**
	2096133	**12**	**PF3D7_1451200**	**conserved Plasmodium protein, unknown function**	**N**	**G908R**	**1.45E-09**
	2098642	**7**	**PF3D7_1451200**	**conserved Plasmodium protein, unknown function**	**S**	**71N**	**2.90E-12**
	2470644	**6**	**PF3D7_1460500**	**conserved Plasmodium protein, unknown function**	**S**	**1518T**	**5.50E-14**
	2481070	**2**	**PF3D7_1460900.1**	**apicoplast ribosomal protein S10 precursor, putative**	**N**	**V127M**	**1.30E-20**
	2488298	45	PF3D7_1461100	conserved Plasmodium protein, unknown function	N	T719R	7.46E-06
	2524926	38	PF3D7_1462300	conserved Plasmodium protein, unknown function	N	D1357N	1.76E-06

Supplementary Table 3 – *Kelch13* alleles observed in the full dataset. The table shows the number of samples carrying each of the 33 *kelch13* non-synonymous mutations (WT=wild type) in our dataset, after excluding samples for which no allele could be determined and heterozygous samples (mixtures of mutant and wild-type parasites). The columns show the kelch allele; sample counts for D. R. Congo, Nigeria, Bangladesh, Myanmar, Thailand, Laos, Cambodia and Vietnam; total sample count; number of samples with phenotype; and mean half-life computed from those phenotypes. Mean HL estimates based on few samples (< 5), which should be considered weakly supported, are shown in italics). Mutations in the upstream *P. falciparum*-specific portion of the gene are shaded violet, and those in the downstream BTB/POZ and propeller domains are shaded blue.

Mutation	Sample Count										Mean HL
	CD	NG	BD	MM	TH	LA	KH	VN	Total	Phenotypes	
WT	64	2	67	47	118	118	292	138	846	630	2.77
K92N	1								1	1	*2.40*
E130G							2		2	2	*2.60*
T149S	2								2	2	*1.74*
K189T	10	2	5	3					20	18	2.39
E252Q				11	8				19	15	3.38
R255K	2								2	2	*2.12*
E270K							1		1	1	*4.59*
D281V				2			1		3	2	*3.33*
D353Y								5	5	5	7.70
F395Y							1		1	1	*3.78*
K438N				1					1	1	*1.38*
P441L				5	6				11	10	6.34
P443S					1				1		-
F446I				3					3	3	*5.02*
G449A				2	3		2		7	7	6.55
N458Y					6				6	6	8.38
A481V					1		2		3	3	*6.13*
Y493H							45	4	49	37	6.76
N525D					1				1	1	*4.68*
N537I					1				1	1	*5.02*
G538V					8				8	8	4.42
R539T					13	2	25	4	44	38	5.70
I543T							2	22	24	23	7.07
P553L					2			9	11	9	6.03
R561H				2	5				7	5	6.93
V568G								5	5	5	6.67
P574L				6	1				7	7	6.85
C580Y				11	21		241	9	282	246	6.72
D584V							2		2	1	*5.41*
F614L					1				1	1	*2.50*
F673I				2					2	2	6.32
A675V				2	11				13	13	5.60
H719N							1		1	1	*5.80*
Total	79	4	72	97	207	120	617	196	1392	1107	

Nature Genetics: doi:10.1038/ng.3189

Supplementary Table 4 – List of single-nucleotide polymorphisms (SNPs) associated with parasite clearance half-life ($PCt_{1/2}$), after correcting for presence of *kelch13* mutations. This GWAS was performed using the same method as for the results in Supplementary Table 2, but the presence of *kelch13* mutations was corrected as a covariate. The 20 most significant associations are shown; polymorphisms that were significant in the initial GWAS are highlighted by a green background. Although several top-scoring SNPs appeared to confer strong phenotypic changes, in most cases there were too few mutant parasites to carry out detailed analyses of these loci. For each SNP, we show chromosome number; nucleotide position; gene id and description; mutation name; whether the SNP is non-synonymous or synonymous; the association *P*-value; and the residual effect of the SNP on $PCt_{1/2}$, after correcting for the *kelch13* mutation effect.

Chr	Pos	Gene Id	Gene Description	Mutation	N/S	*P*	Effect (h)
8	392003	PF3D7_0807600	conserved Plasmodium protein, unknown function	N578D	N	2E-07	1.46
12	2119411	PF3D7_1252100	rhoptry neck protein 3+(RON3)	E691D	N	1E-06	1.85
14	1685702	PF3D7_1441300	serine/threonine protein kinase, putative	1097L	S	4E-06	2.13
12	1114493	PF3D7_1227500	cyclin (CYC2)	2148N	S	5E-06	1.55
12	1784838	PF3D7_1241900	Tetratricopeptide repeat protein, putative	153Q	S	1E-05	1.40
12	1669294	PF3D7_1239800	conserved Plasmodium protein, unknown function	D2948E	N	3E-05	0.98
14	2481070	PF3D7_1460900.1	apicoplast ribosomal protein S10 precursor, putative	V127M	N	4E-05	0.58
14	2504612	PF3D7_1461800	conserved Plasmodium protein, unknown function	K155R	N	5E-05	2.04
14	1956225	PF3D7_1447900	multidrug resistance protein 2+(heavy metal transport family) (MDR2)	T484I	N	5E-05	0.54
14	2355747	PF3D7_1457400	conserved Plasmodium protein, unknown function	E1588V	N	5E-05	0.72
12	1684141	PF3D7_1240000	3-hydroxyisobutyryl-coenzyme A+hydrolase, putative	D436G	N	5E-05	1.36
12	137473	PF3D7_1202300	dynein heavy chain, putative	4572T	S	6E-05	0.54
12	1590289	PF3D7_1238300	cell cycle control protein, putative	T374I	N	7E-05	1.20
12	1759062	PF3D7_1241300	conserved Plasmodium protein, unknown function	46C	S	8E-05	1.26
14	1960049	PF3D7_1448000	U3 snoRNA-associated small subunit rRNA processing protein, put.	H1271R	N	8E-05	0.46
12	1424176	PF3D7_1234100	conserved Plasmodium protein, unknown function	D3255H	N	8E-05	1.14
10	490720	PF3D7_1012700	protein phosphatase, putative	V1157L	N	9E-05	0.52
14	2215809	PF3D7_1453900	conserved Plasmodium protein, unknown function	300T	S	9E-05	0.96
6	116243	PF3D7_0602800	JmjC domain containing protein (JmjC2)	54L	S	9E-05	0.60
12	231311	PF3D7_1205300	conserved Plasmodium protein, unknown function	258K	S	1E-04	2.22

Supplementary Table 5 – Subpopulations of *P. falciparum* identified in Vietnam. The table shows the subpopulations identified in Vietnam samples: VN-C represents the "core" population, VN-F01 to VN-F04 are differentiated founder subpopulations, and VN-U comprises samples that could not be classified in one of the five other populations. The columns show the number of samples; the median and interquartile range of half-life (in hours); the number of samples for which parasite clearance half-life was used in this calculation; and the *P*-value that the distribution of phenotypes differs from that of VN-C (by two-sample Wilcoxon test).

Population	Sample Count	Median Half-life (IQR)		Phenotype Count	*P*
VN-C	76	2.5	(1.9-3.3)	64	
VN-F01	21	7.2	(6.2-8.3)	20	$9x10^{-11}$
VN-F02	12	2.3	(1.8-3.2)	12	0.68
VN-F03	9	2.8	(2.2-3.9)	8	0.18
VN-F04	8	6.4	(4.0-7.5)	8	$7x10^{-5}$
VN-U	79	4.4	(2.6-6.6)	72	10^{-6}
Total	**205**			**184**	

Supplementary Table 6 – Differentiation of SNPs between *P. falciparum* subpopulations in Vietnam. This table shows the number of SNPs that are highly differentiated (as defined by $F_{ST} \geq 0.5$) between each pair of populations. Each cell's background colour intensity is proportional to the value therein.

	VN-C	VN-F01	VN-F02	VN-F03	VN-F04	VN-U
VN-C	-	1673	253	718	754	0
VN-F01	1673	-	2298	3093	3259	1479
VN-F02	253	2298	-	1804	1909	216
VN-F03	718	3093	1804	-	2651	717
VN-F04	754	3259	1909	2651	-	682
VN-U	0	1479	216	717	682	-

Supplementary Table 7 – Subpopulations of *P. falciparum* identified in Cambodia. The table shows the subpopulations identified in Cambodia samples: KH-C represents the "core" population, WKH-F01 to WKH-F04 are differentiated founder subpopulations in west Cambodia, NKH-F01 to NKH-F03 are differentiated founder subpopulations in north Cambodia, and VN-U comprises samples that could not be classified in one of the eight other populations. For each subpopulation, the columns show the number of samples stratified by region (NEKH=northeast, NKH=north, and WKH=west) and their total; the median and interquartile range of half-life (in hours); the number of samples for which parasite clearance half-life was used in this calculation; and the *P*-value that the distribution of phenotypes differs from that of KH-C (by two-sample Wilcoxon test).

Population	Sample Count				Median Half-life (IQR)	Phenotype Count	*P*
	NEKH	NKH	WKH	Total			
KH-C	120	1	3	124	3.0 (2.4-3.5)	119	
WKH-F01			54	54	6.2 (5.2-7.2)	47	$2x10^{-21}$
WKH-F02		2	14	16	6.0 (5.0-6.4)	15	$4x10^{-9}$
WKH-F03			34	34	7.0 (5.8-7.9)	33	$1x10^{-17}$
WKH-F04			15	15	6.4 (5.5-7.1)	13	$7x10^{-9}$
NKH-F01		14		14	3.2 (2.3-3.7)	14	0.25
NKH-F02		9		9	5.5 (5.3-8.4)	9	10^{-5}
NKH-F03		11		11	3.0 (2.8-3.9)	11	0.08
KH-U	11	48	179	238	5.3 (3.4-7.0)	206	10^{-21}
Total	**131**	**85**	**299**	**515**		**467**	

Supplementary Table 8 – Differentiation of SNPs between *P. falciparum* subpopulations in Cambodia. The table shows the number of SNPs that are highly differentiated (as defined by $F_{ST} \geq 0.5$) between each pair of populations. Each cell's background colour intensity is proportional to the value therein.

	KH-C	WKH-F01	WKH-F02	WKH-F03	WKH-F04	NKH-F01	NKH-F02	NKH-F03	KH-U
KH-C	-	1431	1412	1192	654	492	1438	1276	4
WKH-F01	1431	-	3098	1490	2113	2428	3038	3329	644
WKH-F02	1412	3098	-	2568	2174	2385	2997	3239	947
WKH-F03	1192	1490	2568	-	1772	2073	2516	2698	631
WKH-F04	654	2113	2174	1772	-	1596	2145	2364	464
NKH-F01	492	2428	2385	2073	1596	-	2461	2592	498
NKH-F02	1438	3038	2997	2516	2145	2461	-	3270	1172
NKH-F03	1276	3329	3239	2698	2364	2592	3270	-	1202
KH-U	4	644	947	631	464	498	1172	1202	-

Supplementary Table 9 – List of SNPs that are most highly differentiated between founder and core populations in Vietnam and Cambodia. Each SNP is scored by the mean F_{ST} between artemisinin resistant founder populations and their respective core populations. The top 50 results are listed here. SNPs that scored highly in the genome-wide association study ($P \leq 10^{-5}$, Supplementary Table 2) have a coloured background. For each SNP, we show chromosome number; nucleotide position; the containing locus, as identified in Table 1; id and description of the containing gene; whether the SNP is non-synonymous or synonymous; mutation name; non-reference allele frequency (NRAF) in each population; and the score. Headers are shaded pink for artemisinin resistant populations, and blue for sensitive populations.

Chr	Pos	Locus	Gene Id	Gene Description	N/S	Mutation	VN-C	KH-C	VN-F01	VN-F04	WKH-F01	WKH-F02	WKH-F03	WKH-F04	NKH-F02	VN-F02	VN-F03	NKH-F01	NKH-F03	VN-U	KH-U	Score
13	748395	*fd*	PF3D7_1318100	ferredoxin, putative	N	D193Y	0.06	0.02	1.00	0.82	0.98	1.00	1.00	1.00	0.98	0.28	0.00	0.03	0.01	0.51	0.80	0.89
13	754133	*fd*	PF3D7_1318300	conserved Pf protein, unknown function	N	T75I	0.09	0.03	1.00	0.86	0.96	1.00	1.00	1.00	0.98	0.28	0.00	0.02	0.00	0.57	0.82	0.85
7	405600	*crt*	PF3D7_0709000	chloroquine resistance transporter (CRT)	N	I356T	0.15	0.05	1.00	0.00	1.00	1.00	1.00	1.00	0.98	0.33	0.96	0.02	0.04	0.32	0.81	0.75
7	405362	*crt*	PF3D7_0709000	chloroquine resistance transporter (CRT)	N	N326S	0.14	0.06	1.00	0.00	1.00	1.00	1.00	1.00	0.98	0.31	0.94	0.04	0.01	0.31	0.81	0.75
10	497461	*pph*	PF3D7_1012900	conserved Pf protein, unknown function	N	T38I	0.29	0.08	1.00	0.92	1.00	1.00	1.00	1.00	0.97	0.49	0.52	1.00	0.06	0.66	0.77	0.74
5	852479		PF3D7_0520800	conserved Pf protein, unknown function	N	S1183N	0.05	0.03	0.00	0.90	0.96	0.88	1.00	1.00	0.97	0.19	0.01	0.01	0.00	0.35	0.59	0.73
14	2481070	*arps10*	PF3D7_1460900.1	apicoplast ribosomal protein S10 precursor, putative	N	V127M	0.13	0.08	1.00	0.07	1.00	1.00	1.00	1.00	0.98	0.10	0.00	0.02	0.97	0.55	0.77	0.71
14	1481954		PF3D7_1436300	translocon component PTEX150 (PTEX150)	S	726Q	0.08	0.10	1.00	0.92	1.00	0.99	1.00	0.33	0.97	0.01	0.06	0.02	0.04	0.28	0.64	0.70
10	490720	*pph*	PF3D7_1012700	protein phosphatase, putative	N	V1157L	0.31	0.09	1.00	0.92	1.00	1.00	1.00	0.80	0.99	0.48	0.51	1.00	0.07	0.56	0.64	0.68
10	490648	*pph*	PF3D7_1012700	protein phosphatase, putative	N	Y1133N	0.31	0.08	1.00	0.94	1.00	1.00	1.00	0.80	0.97	0.56	0.50	1.00	0.07	0.59	0.65	0.68
5	958440		PF3D7_0523000	multidrug resistance protein (MDR1)	N	Y184F	0.11	0.02	0.00	0.89	1.00	1.00	1.00	0.53	0.98	0.19	0.22	0.02	0.91	0.33	0.61	0.68
12	895047		PF3D7_1222400	transcription factor with AP2 domain(s) (ApiAP2)	N	Q1489H	0.14	0.04	0.00	0.15	1.00	1.00	1.00	1.00	1.00	0.80	0.00	0.02	0.95	0.34	0.60	0.67
5	937428		PF3D7_0522500	mitochondrial ribosomal protein L17 precursor, putative	S	66L	0.10	0.02	0.00	0.86	0.96	1.00	1.00	0.53	0.98	0.19	0.00	0.02	0.92	0.30	0.68	0.67
12	722310		PF3D7_1218400	triose or hexose phosphate/phosphate translocator, putative	N	R294C	0.10	0.02	1.00	0.00	1.00	0.01	1.00	1.00	0.97	0.14	0.00	0.00	0.01	0.31	0.52	0.66
12	895415		PF3D7_1222400	transcription factor with AP2 domain(s) (ApiAP2)	N	E1612V	0.85	0.87	0.00	0.82	0.00	0.00	0.00	0.00	0.00	0.18	0.17	0.97	0.04	0.59	0.38	0.66
5	923033		PF3D7_0522400	conserved Pf protein, unknown function	N	S4938A	0.06	0.05	0.00	0.93	1.00	1.00	1.00	0.53	0.98	0.18	0.00	0.02	0.88	0.25	0.67	0.66
13	2373333	*kelch*	PF3D7_1359600	conserved Pf protein, unknown function	N	L484F	0.08	0.14	1.00	0.26	0.98	1.00	1.00	1.00	0.97	0.27	0.00	0.02	0.00	0.37	0.63	0.65
1	340448		PF3D7_0108300	conserved Pf protein, unknown function	N	S1094N	0.17	0.08	1.00	0.91	0.94	0.13	1.00	1.00	0.98	0.34	0.01	0.20	0.02	0.40	0.57	0.65
12	1010085		PF3D7_1224800	conserved Pf protein, unknown function	N	N266I	0.15	0.04	0.00	0.16	1.00	0.94	1.00	1.00	0.97	0.71	0.81	0.03	0.04	0.19	0.59	0.64
13	2373848	*kelch*	PF3D7_1359600	conserved Pf protein, unknown function	N	D656G	0.15	0.14	1.00	0.23	1.00	1.00	1.00	1.00	0.97	0.27	0.00	0.03	0.00	0.44	0.70	0.64

8	786758		PF3D7_0817300	asparagine-rich antigen	S	953N	0.11	0.05	1.00	0.50	0.90	1.00	0.98	1.00	0.00	0.01	0.72	0.00	0.00	0.30	0.39	0.63
12	892564		PF3D7_1222400	transcription factor with AP2 domain(s) (ApiAP2)	N	H662N	0.19	0.15	1.00	0.14	1.00	0.99	1.00	1.00	1.00	0.80	0.84	0.03	0.98	0.38	0.63	0.63
14	1481740		PF3D7_1436300	translocon component PTEX150 (PTEX150)	N	D655A	0.16	0.12	1.00	0.88	1.00	1.00	1.00	0.33	0.97	0.67	0.06	0.37	0.08	0.38	0.81	0.63
2	524756		PF3D7_0212700	conserved P. protein, unknown function	N	I430V	0.01	0.05	1.00	0.05	0.98	0.99	0.97	0.00	0.92	0.39	0.00	0.03	0.06	0.24	0.37	0.62
9	1022363		PF3D7_0925400	protein phosphatase-beta	N	T299M	0.03	0.07	1.00	0.23	0.95	1.00	1.00	0.00	0.98	0.70	0.00	0.79	0.05	0.33	0.41	0.62
8	511014		PF3D7_0809900	JmjC domain containing protein (JmjC1)	N	S472L	0.26	0.14	1.00	0.94	0.91	0.94	1.00	1.00	0.89	0.02	0.24	0.71	0.00	0.31	0.60	0.62
12	892467		PF3D7_1222400	transcription factor with AP2 domain(s) (ApiAP2)	S	629K	0.18	0.15	1.00	0.15	1.00	1.00	0.99	1.00	1.00	0.79	0.82	0.03	0.97	0.38	0.63	0.62
14	2528725	*arps10*	PF3D7_1462300	conserved Pf protein, unknown function	S	90V	0.17	0.04	1.00	0.13	1.00	0.13	1.00	1.00	0.98	0.09	0.00	0.03	0.98	0.48	0.65	0.62
13	1862741	*kelch*	PF3D7_1346400	conserved Pf protein, unknown function	N	D3282H	0.12	0.07	0.00	0.00	0.94	1.00	1.00	1.00	0.97	0.08	0.01	0.86	0.02	0.25	0.62	0.62
5	855646		PF3D7_0520800	conserved Pf protein, unknown function	S	127N	0.10	0.03	0.00	0.92	0.91	0.87	1.00	0.47	0.98	0.18	0.01	0.00	0.00	0.27	0.52	0.62
5	924826		PF3D7_0522400	conserved Pf protein, unknown function	N	H5535Q	0.12	0.08	0.00	0.91	1.00	1.00	1.00	0.53	0.99	0.19	0.00	0.01	0.90	0.28	0.67	0.61
14	2525648	*arps10*	PF3D7_1462300	conserved Pf protein, unknown function	N	Y1116F	0.14	0.07	0.99	0.07	1.00	0.13	0.99	1.00	1.00	0.11	0.00	0.02	0.97	0.45	0.58	0.61
5	966809		PF3D7_0523200	conserved Pf protein, unknown function	N	N229K	0.21	0.06	1.00	0.91	0.98	1.00	0.53	0.53	0.98	0.39	0.75	0.01	0.05	0.24	0.41	0.61
14	2081506	14-02	PF3D7_1450800	conserved Pf protein, unknown function	N	G248R	0.09	0.14	1.00	0.00	1.00	0.87	1.00	0.88	1.00	0.01	0.00	0.38	0.94	0.18	0.61	0.61
14	2082099	14-02	PF3D7_1450800	conserved Pf protein, unknown function	N	S50N	0.08	0.13	1.00	0.00	1.00	0.80	1.00	0.88	1.00	0.01	0.00	0.37	0.93	0.13	0.61	0.60
10	885314		PF3D7_1021700	conserved Pf protein, unknown function	N	N3000D	0.19	0.09	1.00	0.11	0.96	1.00	1.00	0.60	0.98	0.06	0.76	0.02	0.01	0.40	0.62	0.60
8	789468		PF3D7_0817300	asparagine-rich antigen	N	K50R	0.28	0.18	1.00	0.83	0.94	1.00	1.00	1.00	0.98	0.78	0.76	0.01	0.00	0.60	0.67	0.60
10	331476		PF3D7_1008100	conserved Pf protein, unknown function	N	F77L	0.85	0.83	0.00	0.07	0.00	0.06	0.00	1.00	0.00	0.44	1.00	0.29	0.04	0.68	0.40	0.60
11	796730		PF3D7_1121100	conserved Pf protein, unknown function	N	N777I	0.16	0.06	1.00	0.03	0.98	0.99	1.00	0.07	0.95	0.29	0.98	0.02	0.00	0.29	0.49	0.59
14	1990395	*mdr2*	PF3D7_1448500	chloroquine resistance marker prot. (CRMP)	N	D1727V	0.30	0.14	1.00	0.00	0.99	1.00	1.00	0.80	0.98	0.26	0.00	0.17	0.96	0.28	0.62	0.59
12	1054566		PF3D7_1226000	conserved Pf protein, unknown function	N	H573N	0.08	0.03	0.00	0.77	1.00	0.93	0.99	1.00	0.00	0.06	0.17	0.02	0.91	0.20	0.41	0.59
8	399899		PF3D7_0807800	proteasome subunit alpha type 5, putative	N	E380.	0.16	0.09	1.00	0.68	0.95	0.93	0.99	1.00	0.00	0.02	0.00	1.00	0.02	0.33	0.58	0.59
8	399898		PF3D7_0807800	proteasome subunit alpha type 5, putative	N	E380V	0.16	0.09	1.00	0.68	0.95	0.93	0.99	1.00	0.00	0.02	0.00	1.00	0.02	0.33	0.59	0.59
13	1867630	*kelch*	PF3D7_1346400	conserved Pf protein, unknown function	N	M4911I	0.86	0.90	1.00	1.00	0.06	0.00	0.00	0.00	0.00	1.00	0.99	0.47	0.97	0.76	0.39	0.59
5	806315		PF3D7_0519500	carbon catabolite repressor protein 4, putative (CCR4)	N	I938T	0.03	0.02	0.00	0.73	0.89	0.00	1.00	1.00	0.97	0.00	0.24	0.02	0.00	0.22	0.36	0.59
8	491465		PF3D7_0809600	peptidase family C50, putative	N	C1973Y	0.11	0.02	0.00	0.91	0.91	0.80	1.00	1.00	0.00	0.00	0.00	0.67	0.00	0.20	0.47	0.58
14	1956225	*mdr2*	PF3D7_1447900	multidrug resistance protein 2+ (heavy metal transport family) (MDR2)	N	T484I	0.28	0.22	1.00	0.89	1.00	1.00	1.00	1.00	0.97	0.64	0.00	0.16	0.96	0.52	0.70	0.58
8	399774		PF3D7_0807800	proteasome subunit alpha type 5, putative	S	421K	0.17	0.11	1.00	0.74	0.95	0.92	1.00	1.00	0.00	0.29	0.00	1.00	0.01	0.33	0.57	0.58
1	487131		PF3D7_0112900	Pf exported protein, unknown function	N	I171T	0.22	0.09	1.00	0.00	0.97	1.00	1.00	0.00	0.96	0.48	0.00	0.04	0.00	0.19	0.35	0.58
9	749285		PF3D7_0918200	organelle ribosomal protein L3 prec, putative	S	199R	0.19	0.11	1.00	0.00	1.00	1.00	1.00	1.00	0.00	0.01	0.05	0.04	0.90	0.12	0.48	0.58

Supplementary Table 10 – List of genes with one or more SNPs that are highly differentiated between founder and core populations in Vietnam and Cambodia. This analysis is similar to the analysis that produced results in Supplementary Table 9, but here we adopt a heteroallelic model, i.e. we test a gene rather than a SNP, considering a mutant to be a sample in which any SNP in the gene has mutated. Only SNPs with $F_{ST} \geq 0.3$ between at least one founder and its respective core population are considered in this analysis. The score is the mean of the F_{ST} values for the seven artemisinin resistant founder populations. Here we show the 20 highest-scoring genes, sorted by descending score. For each gene, we show its id and description; score; number of SNPs in the gene; and F_{ST} for each of the artemisinin resistant founder populations. Further analysis has shown that scores tend to increase with SNP count, such that results are most significant for genes with lower SNP counts (Supplementary Figure 6).

Id	Description	Score	SNP Count	WKH-F01	WKH-F02	WKH-F03	WKH-F04	NKH-F02	VN-F04	VN-F02
PF3D7_1343700	**kelch protein, putative**	**0.900**	**5**	**0.92**	**1.00**	**0.99**	**1.00**	**0.95**	**1.00**	**0.45**
PF3D7_1038400	gametocyte-specific protein (Pf11-1)	0.887	17	0.87	0.75	0.86	0.88	0.92	1.00	0.93
PF3D7_1318100	**ferredoxin, putative**	**0.886**	**1**	**0.92**	**0.95**	**0.95**	**0.96**	**0.92**	**0.90**	**0.60**
PF3D7_0826100	E3 ubiquitin-protein ligase, putative	0.881	20	0.81	0.99	0.92	0.96	0.96	0.97	0.54
PF3D7_1318300	**conserved Plasmodium protein, unknown function**	**0.880**	**8**	**0.89**	**0.93**	**0.98**	**1.00**	**0.95**	**0.82**	**0.59**
PF3D7_1303800	conserved Plasmodium protein, unknown function	0.861	28	0.86	0.94	0.98	0.90	0.85	1.00	0.49
PF3D7_1346400	**conserved Plasmodium protein, unknown function**	**0.837**	**7**	**0.88**	**0.98**	**0.87**	**0.99**	**0.95**	**0.69**	**0.49**
PF3D7_0519500	carbon catabolite repressor protein 4, putative (CCR4)	0.824	7	0.88	0.48	0.96	0.96	0.92	0.96	0.60
PF3D7_0701900	**Plasmodium exported protein, unknown function**	**0.822**	**48**	**0.62**	**0.74**	**0.87**	**1.00**	**0.75**	**1.00**	**0.77**
PF3D7_0926100	protein kinase, putative	0.815	8	0.96	0.91	0.91	0.81	0.90	0.64	0.57
PF3D7_1433400	conserved P. membrane protein, unknown function	0.808	9	0.80	0.85	0.97	0.36	0.99	0.92	0.76
PF3D7_1211200	conserved Plasmodium protein, unknown function	0.807	9	0.87	0.92	0.98	0.22	0.93	0.95	0.78
PF3D7_0809600	**peptidase family C50, putative**	**0.802**	**24**	**0.80**	**0.64**	**0.96**	**0.97**	**0.67**	**0.90**	**0.68**
PF3D7_0113800	DBL containing protein, unknown function	0.795	79	0.95	0.78	0.98	0.76	0.62	0.82	0.65
PF3D7_0903400	DEAD/DEAH box helicase, putative	0.793	23	0.91	0.76	0.72	0.75	0.80	0.77	0.84
PF3D7_1418100	liver specific protein 1, putative (LISP1)	0.792	6	0.84	0.93	0.99	0.30	0.81	0.98	0.69
PF3D7_0713900	conserved Plasmodium protein, unknown function	0.791	33	0.73	1.00	0.98	0.41	0.79	0.97	0.66
PF3D7_0914000	pseudouridylate synthase, putative	0.788	7	0.89	0.82	0.96	0.73	0.68	0.76	0.68
PF3D7_1444100	conserved Plasmodium protein, unknown function	0.785	23	0.99	0.55	0.99	0.43	0.91	0.81	0.82
PF3D7_0207300	serine repeat antigen 8+(SERA8)	0.785	8	0.77	0.95	0.76	0.94	0.74	0.98	0.37

Supplementary Table 11 – Presence of *kelch13* alleles in founder and core populations in Vietnam and Cambodia. The table shows the counts of samples carrying each *kelch13* allele (WT = wild type) observed in artemisinin resistant founder and core populations. While the C580Y mutation is practically fixed in three Cambodia founders, it is absent from Vietnam founders and from two Western Cambodia founders. Nearly all parasites in the core populations carry the wild-type *kelch13* allele. Samples classified as "*Het*" carried a mixture of parasites, including mutants.

Population	WT	Y493H	R539T	I543T	P553L	C580Y	*Het*	Total
VN-C	69	1			2		4	**76**
KH-C	122						2	**124**
VN-F01				20			1	**21**
VN-F04	1				4		3	**8**
WKH-F01	2					49	3	**54**
WKH-F02	1		15					**16**
WKH-F03						32	2	**34**
WKH-F04		15						**15**
NKH-F02						8	1	**9**
Total	**195**	**16**	**15**	**20**	**6**	**89**	**16**	**357**

Nature Genetics: doi:10.1038/ng.3189

Supplementary Table 12 – Frequencies of *kelch13* mutants at the 15 Asian sites surveyed. This table shows the proportion of parasites sampled at each location that carried a *kelch13* resistance mutation. The sites are divided into compartments (BD, WSEA and ESEA) following population structure (Figure 3), and ESEA sites are further divided into high (ESEA-HR), variable (ESEA-VR) and low (ESEA-LR) resistance regions, according to the level of resistance mutations observed.

Compartment	Country (Region)	Location	Mutant Frequency
BD	Bangladesh	Bandarban	0.0%
		Ramu	0.0%
WSEA	Myanmar (Central)	Bago	21.8%
	Thailand (West)	Mae Sot	29.2%
	Myanmar (South)	Kawthaung	47.8%
	Thailand (South)	Ranong	66.7%
ESEA-HR	Thailand (East)	Sisakhet	94.4%
	Cambodia (West)	Pailin	91.2%
		Tasanh	79.2%
		Pursat	70.9%
ESEA-VR	Vietnam	Binh Phuoc	29.4%
	Cambodia (North)	Preah Vihear	20.6%
ESEA-LR	Cambodia (Northeast)	Ratanakiri	2.1%
	Laos	Attapeu	2.4%
		Xepon	0.0%

Supplementary Table 13 – Highly differentiated SNPs between population compartments. Estimates of F_{ST} were obtained for two pairs of compartments: Bangladesh (BD) vs. WSEA, and high-resistance (HR) vs. low-resistance (LR) in ESEA. SNPs were subsequently assigned a score equal to the mean of the two estimated. The top 10 results are shown in this table, ordered by descending mean F_{ST}. For each SNP, we show chromosome number; nucleotide position; gene id and description; whether the SNP is non-synonymous or synonymous; mutation name; F_{ST} estimates for the two pairs of compartments, and their mean.

Chr	Pos	Gene Id	Gene Description	N/S	Mutation	F_{ST}		
						WSEA-BD	HR-LR	Mean
13	748395	PF3D7_1318100	ferredoxin, putative	N	D193Y	0.64	0.85	0.75
13	754133	PF3D7_1318300	conserved Plasmodium protein, unknown function	N	T75I	0.46	0.81	0.63
7	405362	PF3D7_0709000	chloroquine resistance transporter (CRT)	N	N326S	0.53	0.71	0.62
14	2481070	PF3D7_1460900.1	apicoplast ribosomal protein S10 precursor, putative	N	V127M	0.44	0.64	0.54
14	1956225	PF3D7_1447900	multidrug resistance protein 2+(heavy metal transport family) (MDR2)	N	T484I	0.53	0.44	0.48
12	1010085	PF3D7_1224800	conserved Plasmodium protein, unknown function	N	N266I	0.31	0.54	0.42
14	2096133	PF3D7_1451200	conserved Plasmodium protein, unknown function	N	G908R	0.43	0.39	0.41
14	1481740	PF3D7_1436300	translocon component PTEX150 (PTEX150)	N	D655A	0.30	0.51	0.41
10	497461	PF3D7_1012900	conserved Plasmodium protein, unknown function	N	T38I	0.15	0.64	0.39
7	405600	PF3D7_0709000	chloroquine resistance transporter (CRT)	N	I356T	0.08	0.70	0.39

Supplementary Table 14 – Frequency of artemisinin resistance mutations in different population compartments. This table uses the same site grouping as in Supplementary Table 12.

Mutation	BD	WSEA	ESEA		
			HR	IR	LR
kelch13	0%	33%	79%	27%	2%
arps10 V127M	0%	61%	92%	42%	12%
fd D193Y	2%	81%	95%	35%	3%
mdr2 T484I	6%	78%	88%	46%	23%
crt N326S	31%	100%	94%	38%	10%

Supplementary Table 15 – Distribution of *kelch13* alleles in the two major Southeast Asian compartments. The table shows, for each *kelch13* resistance mutation observed in our dataset, the number of samples in which the mutant allele is observed, and the number of sites where these samples were found. The alleles are organized in three groups: alleles found only in WSEA, alleles found in both compartments, and alleles found only in ESEA; within each group, alleles are ordered by amino acid position. Only four of 25 alleles were found in both compartments; of these, only C580Y was observed in more than two samples in each compartment.

Allele	WSEA (4 sites)		ESEA (9 sites)	
	Samples	Sites	Samples	Sites
K438N	1	1		
P441L	11	4		
P443S	1	1		
F446I	3	1		
N458Y	6	1		
N525D	1	1		
N537I	1	1		
G538V	8	1		
R561H	7	2		
P574L	7	3		
F614L	1	1		
F673I	2	1		
A675V	13	2		
G449A	5	2	2	1
A481V	1	1	2	1
P553L	2	1	9	1
C580Y	28	3	254	7
D353Y			5	1
F395Y			1	1
Y493H			49	5
R539T			44	7
I543T			24	2
V568G			5	1
D584V			2	2
H719N			1	1
Total	**98**		**398**	

References

1 Takala-Harrison, S. *et al.* Genetic loci associated with delayed clearance of Plasmodium falciparum following artemisinin treatment in Southeast Asia. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 240-245, doi:10.1073/pnas.1211205110 (2013).

2 Bethell, D. *et al.* Artesunate dose escalation for the treatment of uncomplicated malaria in a region of reported artemisinin resistance: a randomized clinical trial. *PLoS One* **6**, e19283, doi:10.1371/journal.pone.0019283 (2011).

3 Hien, T. T. *et al.* In vivo susceptibility of Plasmodium falciparum to artesunate in Binh Phuoc Province, Vietnam. *Malaria journal* **11**, 355, doi:10.1186/1475-2875-11-355 (2012).

4 Mayxay, M. *et al.* No evidence for spread of Plasmodium falciparum artemisinin resistance to Savannakhet Province, Southern Laos. *The American journal of tropical medicine and hygiene* **86**, 403-408, doi:10.4269/ajtmh.2012.11-0497 (2012).

5 Kyaw, M. P. *et al.* Reduced susceptibility of Plasmodium falciparum to artesunate in southern Myanmar. *PLoS One* **8**, e57689, doi:10.1371/journal.pone.0057689 (2013).

6 Dondorp, A. M. *et al.* Artemisinin resistance in Plasmodium falciparum malaria. *N Engl J Med* **361**, 455-467 (2009).

7 Phyo, A. P. *et al.* Emergence of artemisinin-resistant malaria on the western border of Thailand: a longitudinal study. *Lancet*, doi:10.1016/S0140-6736(12)60484-X (2012).

8 Cheeseman, I. H. *et al.* A major genome region underlying artemisinin resistance in malaria. *Science* **336**, 79-82, doi:10.1126/science.1215966 (2012).

9 Amaratunga, C. *et al.* Artemisinin-resistant Plasmodium falciparum in Pursat province, western Cambodia: a parasite clearance rate study. *The Lancet infectious diseases* **12**, 851-858, doi:10.1016/S1473-3099(12)70181-0 (2012).

10 Miotto, O. *et al.* Multiple populations of artemisinin-resistant Plasmodium falciparum in Cambodia. *Nature Genetics*, doi:10.1038/ng.2624 (2013).

11 Ashley, E. A. *et al.* Spread of artemisinin resistance in Plasmodium falciparum malaria. *The New England Journal of Medicine* **371**, 411-423, doi:10.1056/NEJMoa1314981 (2014).

Supplementary Note

Identification of founder populations in Vietnam and Cambodia

To identify populations in Vietnam, we analysed a set of 288 samples: 205 from Vietnam (VN), 48 from northeast Cambodia (NEKH) and 35 from Laos (LA). The Cambodia and Laos parasites were included in the analysis as populations of parasites genetically similar to those in Vietnam, but generally sensitive to artemisinin. The ADMIXTURE-based method (described in Methods) was run for $K = 3 .. 8$, and the resulting labels were used for group assignment, leading to the identification of seven clusters, labelled VN-C and VN-F01 to VN-F06, as well as an "unclassified" group VN-U (Supplementary Figure 11). Cluster VN-C was deemed to represent a core population, due to its consistent grouping with the NEKH and LA populations. Clusters VN-F05 and VN-F06, observed to be unstable in their aggregation with groups of different ancestries (Supplementary Figure 12), were discarded and their samples reassigned to VN-U. The remaining populations (VN-F01 to VN-F04) were tested for evidence of founder effects. All candidate populations had high numbers of differentiated SNPs when compared to VN-C, although this number was considerably lower in VN-F02 ($n = 253$) than in other populations (Supplementary Table 6). All clusters were also found to clearly correspond to groups of PCoA outliers (Supplementary Figure 4).

In Cambodia, the number of populations was expected to be high since at least four populations were previously identified in West and Northeast Cambodia, and samples from a new sampling region of the country (Preah Vihear Province, North Cambodia) were analysed for the first time in this study. Accordingly, ADMIXTURE was run for $K = 3 .. 16$. However, we found that at all $K \geq 12$, newly identified groups either were too small or emerged in the wild-type population without marked differentiation in PCoA plots. Hence, we performed group assignments using ancestry results for $K = 3 .. 11$, identifying a core population KH-C (largely consisting of samples from Northeast Cambodia included in the previously identified KH1 core population[1]), seven candidate founders populations in west Cambodia (WKH-F01 to WKH-F07) and three in north Cambodia (NKH-F01 to NKH-F03), as well as an "unclassified" group KH-U (Supplementary Figure 13). The smaller clusters KH-F06 and KH-F07 comprised samples unclassified for $K \leq 6$, as well as samples that grouped with KH-F03 at $K = 6$ (Supplementary Figure 14). In addition, PCoA plots showed that these two clusters did not separate clearly from the core population in the first nine principal components (data not shown). Hence, KH-F06 and KH-F07 were disregarded, and their samples reassigned to KH-U. PCoA also showed that the KH-F05 cluster did not separate along any principal component, and

F_{ST} analysis showed a very low number of highly differentiated SNPs between this cluster and KH-C (*n* = 49), suggesting that KH-F05 may represent a core population, or an admixed derivative. Thus, KH-F05 was also excluded from the current analysis, and its samples added to VN-U. All remaining candidate founder populations (WKH-F01 to WKH-F04, and NKH-F01 to NKH-F03) were observed to separate from KH-C in PCoA plots (Supplementary Figure 5), and therefore retained for analysis.

References

1 Miotto, O. *et al.* Multiple populations of artemisinin-resistant Plasmodium falciparum in Cambodia. *Nature Genetics*, doi:10.1038/ng.2624 (2013).